

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: John E. Piletz

Tina R. Ivanov

(ii) TITLE OF INVENTION: DNA Sequence Encoding a Human  
Imidazoline Receptor

(iii) NUMBER OF SEQUENCES: 22

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: WENDEROTH, LIND & PONACK

(B) STREET: 805 Fifteenth St. N.W., Suite 700

(C) CITY: Washington

(D) STATE: District of Columbia

(E) COUNTRY: U.S.A.

(F) ZIP: 20005

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.50" 1.44 Mb diskette

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: MS-DOS

(D) SOFTWARE: Wordperfect 5.1+

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: new

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NO.: 60/012,600

(B) FILING DATE: March 1, 1996

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NO.: 08/650,766

(B) FILING DATE: May 20, 1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Warren Cheek

(B) REGISTRATION NUMBER: 33,367

(C) REFERENCE/DOCKET NUMBER: WMC-1342/clone

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (202) 371-8850

(B) TELEFAX: (202) 371-8856

(2) INFORMATION FOR SEQ ID NO: 1

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3389 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: cDNA

(B) CLONE: EST04033 (HFBDP28)

(ix) FEATURE:

(A) NAME/KEY: predicted translation product when  
transfected

(B) LOCATION: 1398 ... 3389

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1

CTCTAGAAC TAGTGGATCC CCCGGGCTGC AGGAATTCCA GTTTAATACT AACCCCTAATG	60
GTGACTGCG GTTTACAAAG AGCTCTGTAT CACCTGGGAT AGCTTTCAGT AGCAATTCAC	120
ACAACCTGGT CCTAAAAAAT AATAACAATA ATAATAATAA TTAGAGAATT AAAACCCAAC	180
TCATGTTGA ATGGTTAAAA TCACGTAAGA ACTGAAATTT GGGGTGGGGG TGTCCTCAAC	240
CTGAGCTT GTCCTAGCAG TGAAAATGCT CGCCTCCAAG CAGGGCTCAG AAAGGTCTGG	300
CCCCTCCAG GCAGAGGGCT GAGCTCAGGG GGCTCTTGGA GGACACTCAC CCCATGGTCC	360
FGGGATGCT TCTGGCTTCC TTAAAAACAG TTGGGCATCC GCATTGTATA AGTAGGTGGA	420
ACCCTAGTG TGGTTCTTTT GAAGGATATG GGAAGGGAGG ATGACGAACT AGAGAAGTGG	480
AGGGGACCA AAATCACTGA GGTCCCAGAA TATCATAGAT TTGGGTATAG GATTGGGGTC	540
TAAGAATT GAGCACCAGG AATTCCAGCT TCTTCCCATT AAAGAAACTG GGAAGTGGTT	600
CCCTTGGAG GCCTATGTAG TGTTTTCTGC CCCTGTCCCA TACCAAGTCT CATTGATATT	660
TCACAGAAT ATCAGATGAA AATCTATTTT TAAAGACCAT TGGGAGAATG GGTGGTGGAG	720
AGGAGTTGG AGTGGGGTTG GGGGGCAGTT AAAAATGAAT AAAAATCTCT CAGCTACAGA	780
CCAAACAT CACTTCCCTC CGCATTCACT GCATTTCCCA GCAGTCCCCA GATGGTTGTT	840
CGTGGGGA CACAGCAGCT GCCTCATTTT CCTTCAGGCC CCATGGGCTG CTGGTCAACC	900
TAGGATCTA CTAAAGATGA CGCAAATGCC GACTGAACAA TCTGAAACCC AAAGGACTCG	960
TGAGAGACA TGTTCTGCTG AGGAGAGAAA GGTGAGCCAA GGGCAGGGCC CAGGTCCCCC	1020
TGGGCCCC CGAGAGCCCC GACATGCACC TTCTGGATGT GTTTGTTCAA GTAGGACTTA	1080
AGCGGAAGA AGCTCCCACA TTCAGGGCAT GGGTACTTCT TCTCCCCATC AGACTCCATT	1140
TGTTTTTG GGACTGCCAT GTCGCAGGAG AAAGAGCCAT TGGCACTCTG CTTCTCTGGC	1200
CTTCAGGT CGCTGGCATC TGAGAGGTCA CCATAGGAGT CAGAGCTCTC AATCGGATCC	1260
TATGTGAGC ATTTCTGGCC TTCTCGGTTA CAGATACTGC AGAAGTTGCT GGGCCCCTCG	1320
TGTGCTTCT TCAGGTGGTC TGCCATGTAT GCTGCCCGCA AGTACTTCCC ACACACCTGG	1380
TGGGCACCT TGTCTTC ATG ACA GGC CAG GTG GGA GCG CAG ACG GTC TCG	1430

Met Thr Gly Gln Val Gly Ala Gln Thr Val Ser

1

5

10

ST GGC AAA AGA AGC ATT GCA GGT CTG ACA CTT GTG AGG CCG CTC AGA 1478

ly Gly Lys Arg Ser Ile Ala Gly Leu Thr Leu Val Arg Pro Leu Arg

15

20

25

ST GTG CAC CTG CTT GAT ATG TCC GTT CAA GTG ATC AGG CCT GGA GAA 1526

er Val His Leu Leu Asp Met Ser Val Gln Val Ile Arg Pro Gly Glu

30

35

40

DC TTT CCC ACA GCT CTG GCA GAT GTA AGG CGG AAT TCC CCA GAG AAG 1574

la Phe Pro Thr Ala Leu Ala Asp Val Arg Trp Asn Ser Pro Glu Lys

45

50

55

AG GGT GGT GAA GAC TCC CGG CTC TCA GCT GCC CCC TGC ATC AGA CCC 1622

ys Gly Gly Glu Asp Ser Trp Leu Ser Ala Ala Pro Cys Ile Arg Pro

65

70

75

SC AGC TCC CCT CCC ACT GTG GCT CCC GCA TCT GCC TCC CTG CCC CAG 1670

er Ser Ser Pro Pro Thr Val Ala Pro Ala Ser Ala Ser Leu Pro Gln

80

85

90

DC ATC CTC TCT AAC CAA GGA ATC ATG TTC GTT CAG GAG GAG GCC CTG 1718

to Ile Leu Ser Asn Gln Gly Ile Met Phe Val Gln Glu Glu Ala Leu

95

100

105

57

CC AGC AGC CTC TCG TCC ACT GAC AGT CTG ACT CCC GAG CAC CAG CCC 1766  
 la Ser Ser Leu Ser Ser Thr Asp Ser Leu Thr Pro Glu His Gln Pro  
 110 115 120

TT GCC CAG GGA TGT TCT GAT TCC TTG GAG TCC ATC CCT GCG GGA CAG 1814  
 le Ala Gln Gly Cys Ser Asp Ser Leu Glu Ser Ile Pro Ala Gly Gln  
 125 130 135

CA GCT TCC GAT GAT TTA AGG GAC GTG CCA GGA GCT GTT GGT GGT GCA 1862  
 la Ala Ser Asp Asp Leu Arg Asp Val Pro Gly Ala Val Gly Gly Ala  
 40 145 150 155

GC CCA GAA CAT GCC GAG CCG GAG GTC CAG GTG GTG CCG GGG TCT GGC 1910  
 er Pro Glu His Ala Glu Pro Glu Val Gln Val Val Pro Gly Ser Gly  
 160 165 170

AG ATC ATC TTC CTG CCC TTC ACC TGC ATT GGC TAC ACG GCC ACC AAT 1958  
 ln Ile Ile Phe Leu Pro Phe Thr Cys Ile Gly Tyr Thr Ala Thr Asn  
 175 180 185

AG GAC TTC ATC CAG CGC CTG AGC ACA CTG ATC CGG CAG GCC ATC GAG 2006  
 ln Asp Phe Ile Gln Arg Leu Ser Thr Leu Ile Trp Gln Ala Ile Glu  
 190 195 200

CG CAG CTG CCT GCC TGG ATC GAG GCT GCC AAC CAG CGG GAG GAG GGC 2054  
 rp Gln Leu Pro Ala Trp Ile Glu Ala Ala Asn Gln Trp Glu Glu Gly

205

210

215

AG GGT GAA CAG GGC GAG GAG GAG GAT GAG GAG GAG GAA GAA GAG GAG 2102

ln Gly Glu Gln Gly Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Glu

20

225

230

235

AC GTG GCT GAG AAC CGC TAC TTT GAA ATG GGG CCC CCA GAC GTG GAG 2150

sp Val Ala Glu Asn Arg Tyr Phe Glu Met Gly Pro Pro Asp Val Glu

240

245

250

AG GAG GAG GGA GGA GGC CAG GGG GAG GAA GAG GAG GAG GAA GAG GAG 2198

lu Glu Glu Gly Gly Gly Gln Gly Glu Glu Glu Glu Glu Glu Glu Glu

255

260

265

AT GAA GAG GCC GAG GAG GAG CGC CTG GCT CTG GAA TGG GCC CTG GGC 2246

sp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu Glu Trp Ala Leu Gly

270

275

280

CG GAC GAG GAC TTC CTG CTG GAG CAC ATC CGC ATC CTC AAG GTG CTG 2294

la Asp Glu Asp Phe Leu Leu Glu His Ile Arg Ile Leu Lys Val Leu

285

290

295

GG TGC TTC CTG ATC CAT GTG CAG GGC AGT ATC CGC CAG TTC GCC GCC 2342

rp Cys Phe Leu Ile His Val Gln Gly Ser Ile Arg Gln Phe Ala Ala

300

305

310

315

59

GC CTT GTG CTC ACC GAC TTC GGC ATC GCA GTC TTC GAG ATC CCG CAC 2390  
ys Leu Val Leu Thr Asp Phe Gly Ile Ala Val Phe Glu Ile Pro His

320

325

330

AG GAG TCT CGG GGC AGC AGC CAG CAC ATC CTC TCC TCC CTG CGC TTT 2438  
ln Glu Ser Trp Gly Ser Ser Gln His Ile Leu Ser Ser Leu Arg Phe

335

340

345

TC TTT TGC TTC CCG CAT GGC GAC CTC ACC GAG TTT GGC TTC CTC ATG 2486  
al Phe Cys Phe Pro His Gly Asp Leu Thr Glu Phe Gly Phe Leu Met

350

355

360

CG GAG CTG TGT CTG GTG CTC AAG GTA CGG CAC AGT GAG AAC ACG CTC 2534  
ro Glu Leu Cys Leu Val Leu Lys Val Arg His Ser Glu Asn Thr Leu

365

370

375

TC ATT ATC TCG GAC GCC GCC AAC CTG CAC GAG TTC CAC GCG GAC CTG 2582  
he Ile Ile Ser Asp Ala Ala Asn Leu His Glu Phe His Ala Asp Leu

80

385

390

395

GC TCA TGC TTT GCA CCC CAG CAC ATG GCC ATG CTG TGT AGC CCC ATC 2630  
rg Ser Cys Phe Ala Pro Gln His Met Ala Met Leu Cys Ser Pro Ile

400

405

410

TC TAC GGC AGC CAC ACC AGC CTG CAG GAG TTC CTG CGC CAG CTG CTC 2678  
eu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe Leu Arg Gln Leu Leu

60

415

420

425

CC TTC TAC AAG GTG GCT GGC GGC TGC CAG GAG CGC AGC CAG GGC TGC 2726  
ar Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu Arg Ser Gln Gly Cys  
430 435 440

TC CCC GTC TAC CTG GTC TAC AGT GAC AAG CGC ATG GTG CAG ACG GCC 2774  
ne Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala  
445 450 455

CC GGG GAC TAC TCA GGC AAC ATC GAG TGG GCC AGC TGC ACA CTC TGT 2822  
la Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys  
50 465 470 475

CA GCC GTG CGG CGC TCC TGC TGC GCG CCC TCT GAG GCC GTC AAG TCC 2870  
ar Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser  
480 485 490

CC GCC ATC CCC TAC TGG CTG TTG CTC ACG CCC CAG CAC CTC AAC GTC 2918  
la Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val  
495 500 505

CC AAG GCC GAC TTC AAC CCC ATG CCC AAC CGT GGC ACC CAC AAC TGT 2966  
e Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys  
510 515 520

61



IC AAC CGC AAC AGC TTC AAG CTC AGC CGT GTG CCG CTC TCC ACC GTG 3014  
 rg Asn Arg Asn Ser PHe Lys Leu Ser Arg Val Pro Leu Ser Thr Val  
 525 530 535

IG CTG GAC CCC ACA CGC AGC TGT ACC CAG CCT CGG GGC GCC TTT GCT 3062  
 u Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala  
 545 550 555

AT GGC CAC GTG CTA GAG CTG CTC GTG GGG TAC CGC TTT GTC ACT GCC 3110  
 p Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala  
 560 565 570

IC TTC GTG CTG CCC CAC GAG AAG TTC CAC TTC CTG CGC GTC TAC AAC 3158  
 e Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn  
 575 580 585

IG CTG CGG GCC TCG CTG CAG GAC CTG AAG ACT GTG GTC ATC GCC AAG 3206  
 n Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys  
 590 595 600

IC CCC GGG ACG GGA GGC AGC CCC CAG GGC TCC TTT GCG GAT GGC CAG 3254  
 r Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln  
 605 610 615

T GCC GAG CGC AGG GCC AGC AAT GAC CAG CGT CCC CAG GAG GTC CCA 3302  
 o Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro

62

20

625

630

635

CA GAG GCT CTG GCC CCG GCC CCA GTG GAA GTC CCA GCT CCA GCC CCG 3350

la Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro

640

645

650

AA TTC GAT ATC AAG CTT ATC GAT ACC GTC GAC CTG CAG

3389

lu Phe Asp Ile Lys Leu Ile Asp Thr Val Asp Leu Gln

655

660

664

## 2) INFORMATION FOR SEQ ID NO: 2

### (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1954 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

### (ix) SEQUENCE DESCRIPTION: SEQ ID NO: 2

TGACAGGCC AGGTGGGAGC GCAGACGGTC TCGGGTGGCA AAAGAAGCAT TGCAGGTCTG 60

CAGTTGTGA GGCCGCTCAG AAGTGTGCAC CTGCTTGATA TGTCCGTTCA AGTGATCAGG 120

CTGGAGAAG CCTTTCCCAC AGCTCTGGCA GATGTAAGGC GGAATTCCCC AGAGAAGAAG 180

GTGGTGAAG ACTCCCGGCT CTCAGCTGCC CCCTGCATCA GACCCAGCAG CTCCCCTCCC 240

CTGTGGCTC CCGCATCTGC CTCCCTGCCC CAGCCCATCC TCTCTAACCA AGGAATCATG 300

TCGTTCAGG AGGAGGCCCT GGCCAGCAGC CTCTCGTCCA CTGACAGTCT GACTCCCGAG 360

ACCAGCCCA TTGCCAGGG ATGTTCTGAT TCCTTGAGT CCATCCCTGC GGGACAGGCA 420

CTTCCGATG ATTTAAGGGA CGTGCCAGGA GCTGTTGGTG GTGCAAGCCC AGAACATGCC 480

AGCCGGAGG TCCAGGTGGT GCCGGGGTCT GGCCAGATCA TCTTCCTGCC CTTACCTGC 540

63

TTGGCTACA	CGGCCACCAA	TCAGGACTTC	ATCCAGCGCC	TGAGCACACT	GATCCGGCAG	600
CCATCGAGC	GGCAGCTGCC	TGCCTGGATC	GAGGCTGCCA	ACCAGCGGGA	GGAGGGCCAG	660
GTGAACAGG	GCGAGGAGGA	GGATGAGGAG	GAGGAAGAAG	AGGAGGACGT	GGCTGAGAAC	720
GCTACTTTG	AAATGGGGCC	CCCAGACGTG	GAGGAGGAGG	AGGGAGGAGG	CCAGGGGGAG	780
AAGAGGAGG	AGGAAGAGGA	GGATGAAGAG	GCCGAGGAGG	AGCGCCTGGC	TCTGGAATGG	840
CCCTGGGCG	CGGACGAGGA	CTTCCTGCTG	GAGCACATCC	GCATCCTCAA	GGTGCTGTGG	900
GCTTCCTGA	TCCATGTGCA	GGGCAGTATC	CGCCAGTTTCG	CCGCCTGCCT	TGTGCTCACC	960
ACTTCGGCA	TCGCAGTCTT	CGAGATCCCG	CACCAGGAGT	CTCGGGGCAG	CAGCCAGCAC	1020
TCCTCTCCT	CCCTGCGCTT	TGTCTTTTGC	TTCCCGCATG	GCGACCTCAC	CGAGTTTGGC	1080
TCCTCATGC	CGGAGCTGTG	TCTGGTGCTC	AAGGTACGGC	ACAGTGAGAA	CACGCTCTTC	1140
TTATCTCGG	ACGCCGCCAA	CCTGCACGAG	TTCCACGCGG	ACCTGCGCTC	ATGCTTTGCA	1200
CCGAGCACA	TGGCCATGCT	GTGTAGCCCC	ATCCTCTACG	GCAGCCACAC	CAGCCTGCAG	1260
AGTTCCTGC	GCCAGCTGCT	CACCTTCTAC	AAGGTGGCTG	GCGGCTGCCA	GGAGCGCAGC	1320
AGGGCTGCT	TCCCCGTCTA	CCTGGTCTAC	AGTGACAAGC	GCATGGTGCA	GACGGCCGCC	1380
GGGACTACT	CAGGCAACAT	CGAGTGGGCC	AGCTGCACAC	TCTGTTCAGC	CGTGCGGCGC	1440
CCTGCTGCG	CGCCCTCTGA	GGCCGTCAAG	TCCGCCGCCA	TCCCCTACTG	GCTGTTGCTC	1500
CGCCCCAGC	ACCTCAACGT	CATCAAGGCC	GACTTCAACC	CCATGCCCAA	CCGTGGCACC	1560
ACAACTGTC	GCAACCGCAA	CAGCTTCAAG	CTCAGCCGTG	TGCCGCTCTC	CACCGTGCTG	1620
TGGACCCCA	CACGCAGCTG	TACCCAGCCT	CGGGGCGCCT	TTGCTGATGG	CCACGTGCTA	1680
AGCTGCTCG	TGGGGTACCG	CTTTGTCACT	GCCATCTTCG	TGCTGCCCCA	CGAGAAGTTC	1740
ACTTCCTGC	GCGTCTACAA	CCAGCTGCGG	GCCTCGCTGC	AGGACCTGAA	GA CTGTGGTC	1800
TCGCCAAGA	CCCCCGGGAC	GGGAGGCAGC	CCCCAGGGCT	CCTTTGCGGA	TGGCCAGCCT	1860
CCGAGCGCA	GGGCCAGCAA	TGACCAGCGT	CCCCAGGAGG	TCCCAGCAGA	GGCTCTGGCC	1920
CGGCCCCAG	TGGAAGTCCC	AGCTCCAGCC	CCGG			1954

64

3) INFORMATION FOR SEQ ID NO: 3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3318 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 3

ATTCCAGTT TAATACTAAC CCTAATGTGT GACTGCGGTT TACAAAGAGC TCTGTATCAC	60
GTGGGATAGC TTTCAGTAGC AATTCACACTAC AACTGGTCCT AAAAAATAAT AACATAATA	120
TAATAATTA GAGAATTAAC ACCCAACAGC ATGTTGAATG GTTAAATCA CGTAAGAACT	180
AAATTTGGG GTGGGGGTGT CCTCAACAGC TGAGCTTGTC CTAGCAGTGA AAATGCTCGC	240
TCGAAGCAG GGCTCAGAAA GGTCTGGAGC CCTCCAGGCA GAGGGCTGAG CTCAGGGGGC	300
CTTGGAGGA CACTCACCCC ATGGTCCATG GGATGCTTCT GGCTTCCTTA AAAACAGTTG	360
GCATCCGCA TTGTATAAGT AGGTGGAGAC CCTAGTGTGG TTCTTTTGAA GGATATGGGA	420
GGGAGGATG ACGAACTAGA GAAGTGGGAG GGGACCAAAA TCACTGAGGT CCCAGAATAT	480
ATAGATTTG GGTATAGGAT TGGGGTCACT AAGAATTGAG CACCAGGAAT TCCAGCTTCT	540
CCGATTAAA GAAACTGGGA CTGGTTTTGC CTTGGAGGCC TATGTAGTGT TTTCTGCCCC	600
GTCCCATAC CAAGTCTCAT TGATATTTCT GCAGAATATC AGATGAAAAT CTATTTCTAA	660
GACCATTTG GAGAATGGGT GGTGGAGAAG GAGTTGGAGT GGGGTTGGGG GGCAGTTAAA	720
ATGAATAAA AATCTCTCAG CTACAGAACC CAAACATCAC TTCCCTCCGC ATTCACAGCA	780
TTCCCAGCA GTCCCCAGAT GGTGTTTCC GTGGGGACAC AGCAGCTGCC TCATTTCCCT	840
CAGGCCCCA TGGGCTGCTG GTCAACCTCA GGATCTACTA AAGATGACGC AAATGCCGAC	900
GAACAATCT GAAACCCAAA GGAATCGAGG AGAGACATGT TCTGCTGAGG AGAGAAAGGT	960
AGCCAAGGG CAGGGCCCAG GTCCCCCAGG GGGCCCCCGA GAGCCCGGAC ATGCACCTTC	1020
GGATGTGTT TGTTCAAGTA GGAATAGAG CGGAAGAAGC TCCCACATTC AGGGCATGGG	1080
ACTTCTTCT CCCCATCAGA CTCCATTTTG TTTTGGGGA CTGCCATGTC GCAGGAGAAA	1140

AGCCATTGG	CACTCTGCTT	CTCTGGCGTC	TTCAGGTCGC	TGGCATCTGA	GAGGTCACCA	1200
AGGAGTCAG	AGCTCTCAAT	CGGATCCTGA	TGTGAGCATT	TCTGGCCTTC	TCGGTTACAG	1260
FACTGCAGA	AGTTGCTGGG	CCCCTCGCTG	TGCTTCTTCA	GGTGGTCTGC	CATGTATGCT	1320
CCCGCAAGT	ACTTCCCACA	CACCTGGCAG	GGCACCTTGT	CTTCATGACA	GGCCAGGTGG	1380
AGCGCAGAC	GGTCTCGGGT	GGCAAAAGAA	GCATTGCAGG	TCTGACACTT	GTGAGGCCGC	1440
CAGAAGTGT	GCACCTGCTT	GATATGTCCG	TTCAAGTGAT	CAGGCCTGGA	GAAGCCTTTC	1500
CACAGCTCT	GGCAGATGTA	AGGCGGAATT	CCCCAGAGAA	GAAGGGTGGT	GAAGACTCCC	1560
ECTCTCAGC	TGCCCCCTGC	ATCAGACCCA	GCAGCTCCCC	TCCCACTGTG	GCTCCCGCAT	1620
FGCCTCCCT	GCCCCAGCCC	ATCCTCTCTA	ACCAAGGAAT	CATGTTCTGT	CAGGAGGAGG	1680
CTGGCCAG	CAGCCTCTCG	TCCACTGACA	GTCTGACTCC	CGAGCACCAG	CCCATTGCCC	1740
GGATGTTC	TGATTCCTTG	GAGTCCATCC	CTGCGGGACA	GGCAGCTTCC	GATGATTTAA	1800
GGCGTGCC	AGGAGCTGTT	GGTGGTGCAA	GCCCAGAACA	TGCCGAGCCG	GAGGTCCAGG	1860
GTGCCGGG	GTCTGGCCAG	ATCATCTTCC	TGCCCTTCAC	CTGCATTGGC	TACACGGCCA	1920
CAATCAGGA	CTTCATCCAG	CGCCTGAGCA	CACTGATCCG	GCAGGCCATC	GAGCGGCAGC	1980
CGTGCCTG	GATCGAGGCT	GCCAACCAGC	GGGAGGAGGG	CCAGGGTGAA	CAGGGCGAGG	2040
GAGGATGA	GGAGGAGGAA	GAAGAGGAGG	ACGTGGCTGA	GAACCGCTAC	TTTGAAATGG	2100
CGCCCGAGA	CGTGGAGGAG	GAGGAGGGAG	GAGGCCAGGG	GGAGGAAGAG	GAGGAGGAAG	2160
GAGGATGA	AGAGGCCGAG	GAGGAGCGCC	TGGCTCTGGA	ATGGGCCCTG	GGCGCGGACG	2220
GACTTCCT	GCTGGAGCAC	ATCCGCATCC	TCAAGGTGCT	GTGGTGCTTC	CTGATCCATG	2280
CAGGGCAG	TATCCGCCAG	TTCGCCGCCT	GCCTTGCTGCT	CACCGACTTC	GGCATCGCAG	2340
TTTCGAGAT	CCCGCACCAG	GAGTCTCGGG	GCAGCAGCCA	GCACATCCTC	TCCTCCCTGC	2400
TTTGTCTT	TTGCTTCCCC	CATGGCGACC	TCACCGAGTT	TGGCTTCCTC	ATGCCGGAGC	2460
TGTCTGGT	GCTCAAGGTA	CGGCACAGTG	AGAACACGCT	CTTCATTATC	TCGGACGCCG	2520
AACCTGCA	CGAGTTCCAC	GCGGACCTGC	GCTCATGCTT	TGCACCCCAG	CACATGGCCA	2580
CTGTGTAG	CCCCATCCTC	TACGGCAGCC	ACACCAGCCT	GCAGGAGTTC	CTGCGCCAGC	2640
CTCACCTT	CTACAAGGTG	GCTGGCGGCT	GCCAGGAGCG	CAGCCAGGGC	TGCTTCCCCG	2700

66

TACCTGGT	CTACAGTGAC	AAGCGCATGG	TGCAGACGGC	CGCCGGGGAC	TACTCAGGCA	2760
ATCGAGTG	GGCCAGCTGC	ACACTCTGTT	CAGCCGTGCG	GCGCTCCTGC	TGCGCGCCCT	2820
GAGGCCGT	CAAGTCCGCC	GCCATCCCCT	ACTGGCTGTT	GCTCACGCCC	CAGCACCTCA	2880
GTCATCAA	GGCCGACTTC	AACCCCATGC	CCAACCGTGG	CACCCACAAC	TGTCGCAACC	2940
AACAGCTT	CAAGCTCAGC	CGTGTGCCGC	TCTCCACCGT	GCTGCTGGAC	CCCACACGCA	3000
TGTACCCA	GCCTCGGGGC	GCCTTTGCTG	ATGGCCACGT	GCTAGAGCTG	CTCGTGGGGT	3060
CGCTTTGT	CACTGCCATC	TTCGTGCTGC	CCCACGAGAA	GTTCCACTTC	CTGCGCGTCT	3120
AACCAGCT	GCGGGCCTCG	CTGCAGGACC	TGAAGACTGT	GGTCATCGCC	AAGACCCCCG	3180
ACGGGAGG	CAGCCCCCAG	GGCTCCTTTG	CGGATGGCCA	GCCTGCCGAG	CGCAGGGCCA	3240
AATGACCA	GCGTCCCCAG	GAGGTCCCAG	CAGAGGCTCT	GGCCCCGGCC	CCAGTGGAAG	3300
CCAGCTCC	AGCCCCGG					3318

) INFORMATION FOR SEQ ID NO: 4

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 4

GGAGGAGG	AAGAGGAGGA	TGAAGAGGCC	GAGGAGGAGC	GCCTGGCTCT	GGAATGGGCC	60
GGGCGCGG	ACGAGGACTT	CCTGCTGGAG	CACATCCGCA	TCCTCAAGGT	GCTGTGGTGC	120
CTGATCC	ATGTGCAGGG	CAGTATCCGC	CAGTTCGCCG	CCTGCCTTGT	GCTCACCGAC	180
GGCATCG	CAGTCTTCGA	GATCCCGCAC	CAGGAGTCTC	GGGGCAGCAG	CCAGCACATC	240
TCCTCCC	TGCGCTTTGT	CTTTTGCTTC	CCGCATGGCG	ACCTCACCGA	GTTTGGCTTC	300
CATGCCGG	AGCTGTGTCT	GGTGCTCAAG	GTACGGCACA	GTGAGAACAC	GCTCTTCATT	360

67

CTCGGACG CCGCCAACCT GCACGAGTTC CACGCGGACC TGCGCTCATG CTTTGCACCC 420  
 GCACATGG CCATGCTGTG TAGCCCCATC CTCTACGGCA GCCACACCAG CCTGCAGGAG 480  
 CCTGCGCC AGCTGCTCAC CTTCTACAAG GTGGCTGGCG GCTGCCAGGA GCGCAGCCAG 540  
 CTGCTTCC CCGTCTACCT GGTCTACAGT GACAAGCGCA TGGTGCAGAC GGCCGCCGGG 600  
 CTACTCAG GCAACATCGA GTGGGCCAGC TGCACACTCT GTTCAGCCGT GCGGCGCTCC 660  
 CTGCGCGC CCTCTGAGGC CGTCAAGTCC GCCGCCATCC CCTACTGGCT GTTGCTCACG 720  
 CCAGCACC TCAACGTCAT CAAGGCCGAC TTCAACCCCA TGCCCAACCG TGGCACCCAC 780  
 CTGTGCGA ACCGCAACAG CTTCAAGCTC AGCCGTGTGC CGCTCTCCAC CGTGCTGCTG 840  
 CCCCACAC GCAGCTGTAC CCAGCCTCGG GGCGCCTTTG CTGATGGCCA CGTGCTAGAG 900  
 GCTCGTGG GGTACCGCTT TGTCACTGCC ATCTTCGTGC TGCCCCACGA GAAGTTCCAC 960  
 CCTGCGCG TCTACAACCA GCTGCGGGCC TCGCTGCAGG ACCTGAAGAC TGTGGTCATC 1020  
 CAAGACCC CCGGGACGGG AGGCAGCCCC CAGGGCTCCT TTGCGGATGG CCAGCCTGCC 1080  
 GCGCAGGG CCAGCAATGA CCAGCGTCCC CAGGAGGTCC CAGCAGAGGC TCTGGCCCCG 1140  
 CCCAGTGG AAGTCCCAGC TCCAGCCCCG G 1171

(5) INFORMATION FOR SEQ ID NO: 5

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 651 amino acids

(B) TYPE: polypeptide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 5

Met Thr Gly Gln Val Gly Ala Gln Thr Val Ser  
 1 5 10

Gly Gly Lys Arg Ser Ile Ala Gly Leu Thr Leu Val Arg Pro Leu Arg  
 15 20 25

68







Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala  
445 450 455

Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys  
460 465 470 475

Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser  
480 485 490

Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val  
495 500 505

Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys  
510 515 520

Arg Asn Arg Asn Ser Phe Lys Leu Ser Arg Val Pro Leu Ser Thr Val  
525 530 535

Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala  
540 545 550 555

Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala  
560 565 570

Ile Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn  
575 580 585

Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys  
590 595 600

Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln  
605 610 615

Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro  
620 625 630 635

Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro  
640 645 650

71

(6) INFORMATION FOR SEQ ID NO: 6

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 390 amino acids

(B) TYPE: polypeptide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 6

Glu Glu Glu Glu Glu Glu  
1 5

Asp Glu Glu Ala Glu Glu Glu Arg Leu Ala Leu Glu Trp Ala Leu Gly  
10 15 20

Ala Asp Glu Asp Phe Leu Leu Glu His Ile Arg Ile Leu Lys Val Leu  
25 30 35

Trp Cys Phe Leu Ile His Val Gln Gly Ser Ile Arg Gln Phe Ala Ala  
40 45 50

Cys Leu Val Leu Thr Asp Phe Gly Ile Ala Val Phe Glu Ile Pro His  
55 60 65 70

Gln Glu Ser Trp Gly Ser Ser Gln His Ile Leu Ser Ser Leu Arg Phe  
75 80 85

Val Phe Cys Phe Pro His Gly Asp Leu Thr Glu Phe Gly Phe Leu Met  
90 95 100

Pro Glu Leu Cys Leu Val Leu Lys Val Arg His Ser Glu Asn Thr Leu  
105 110 115

Phe Ile Ile Ser Asp Ala Ala Asn Leu His Glu Phe His Ala Asp Leu  
120 125 130

Arg Ser Cys Phe Ala Pro Gln His Met Ala Met Leu Cys Ser Pro Ile  
135 140 145 150

Leu Tyr Gly Ser His Thr Ser Leu Gln Glu Phe Leu Arg Gln Leu Leu

155

160

165

Thr Phe Tyr Lys Val Ala Gly Gly Cys Gln Glu Arg Ser Gln Gly Cys  
170 175 180

Phe Pro Val Tyr Leu Val Tyr Ser Asp Lys Arg Met Val Gln Thr Ala  
185 190 195

Ala Gly Asp Tyr Ser Gly Asn Ile Glu Trp Ala Ser Cys Thr Leu Cys  
200 205 210

Ser Ala Val Arg Arg Ser Cys Cys Ala Pro Ser Glu Ala Val Lys Ser  
215 220 225 230

Ala Ala Ile Pro Tyr Trp Leu Leu Leu Thr Pro Gln His Leu Asn Val  
235 240 245

Ile Lys Ala Asp Phe Asn Pro Met Pro Asn Arg Gly Thr His Asn Cys  
250 255 260

Arg Asn Arg Asn Ser Phe Lys Leu Ser Arg Val Pro Leu Ser Thr Val  
265 270 275

Leu Leu Asp Pro Thr Arg Ser Cys Thr Gln Pro Arg Gly Ala Phe Ala  
280 285 290

Asp Gly His Val Leu Glu Leu Leu Val Gly Tyr Arg Phe Val Thr Ala  
295 300 305 310

Ile Phe Val Leu Pro His Glu Lys Phe His Phe Leu Arg Val Tyr Asn  
315 320 325

Gln Leu Arg Ala Ser Leu Gln Asp Leu Lys Thr Val Val Ile Ala Lys  
330 335 340

Thr Pro Gly Thr Gly Gly Ser Pro Gln Gly Ser Phe Ala Asp Gly Gln  
345 350 355

Pro Ala Glu Arg Arg Ala Ser Asn Asp Gln Arg Pro Gln Glu Val Pro

73

360

365

370

Ala Glu Ala Leu Ala Pro Ala Pro Val Glu Val Pro Ala Pro Ala Pro  
375 380 385 390

(7) INFORMATION FOR SEQ ID NO: 7

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 7

CTTGAGGATG CGGATGTGCT 20

(8) INFORMATION FOR SEQ ID NO: 8

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 8

CCATGGGGTG AGTGCCT 18

(9) INFORMATION FOR SEQ ID NO: 9

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid

74

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 9

AGGACACTCA CCCCATGG 18

(10) INFORMATION FOR SEQ ID NO: 10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 10

GTATGGGACA GGGGCAGAAA 20

(11) INFORMATION FOR SEQ ID NO: 11

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 11

TTTCTAAAGA CCATTGGGAG 20

(12) INFORMATION FOR SEQ ID NO: 12

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 12

CCATTTTAAA GTAGCGGTTC 20

(13) INFORMATION FOR SEQ ID NO: 13

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 13

AGGAGAGAAA GGTGAGCCAA 20

(14) INFORMATION FOR SEQ ID NO: 14

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 14

GTAGATCCTG AGGTTGACCA 20

(15) INFORMATION FOR SEQ ID NO: 15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 15

TGTGAGCATT TCTGGCCTTC 20

(16) INFORMATION FOR SEQ ID NO: 16

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 16

TGAAGACGCC AGAGAAGCAG 20

(17) INFORMATION FOR SEQ ID NO: 17

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 17

GCCTCACAAG TGTCAGACCT 20



(18) INFORMATION FOR SEQ ID NO: 18

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 18

AGAAGGGTGG TGAAGACT 18

(19) INFORMATION FOR SEQ ID NO: 19

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 19

CTTGTTAGA GAGGATGGGC 20

(20) INFORMATION FOR SEQ ID NO: 20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 20

GCCCATCCTC TCTAACCAAG 20

(21) INFORMATION FOR SEQ ID NO: 21

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15202 nucleic acids
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

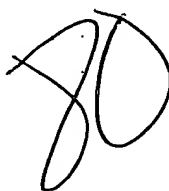
(ix) FEATURE:

- (A) NAME/KEY:
- (B) LOCATION:
- (C) IDENTIFICATION METHOD:
- (D) OTHER INFORMATION: /note="N is unknown or other"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21

GATCCGAGCTCAATTAACCCTCACTAAAGGGAGTCGACTCGATCCTTAAA  
ATATTCATATCTCCTGGACAACCTGTGGCCATAGTGCCTGACTGTAAACC  
CAAAGGGTTTGCTTTGCCAGTGTAGCCCAGCCTGGTGTCTGCTGCCCCCT  
CGCGGTGTCTGTGCACCTGCCACGATGCTGACCAGACACCCTTAACCAGG  
TTCACCCATCGCCTGGGCCTGGAGCAGTCCCCCTGATGCTCTGATTGGTC  
CTTGGACCTTCTGTTCTCCCAAATCCCAGGTCAGAAAATACCTGGAAGT  
CTATTTGTGTCCACCTCCCTCTTTGTGGCCGCAAGTGCCCCTTCCTCCA  
CACAGTCACAAGACCATGAGATGCCATCTCCTCCCCTCCTGGGCTGCAGA  
CTTTGGGAAGCTCCCAGGCCACAGAGGTGTCAGCTCCTGTCCAGGCCCTT  
GGGACCTTCCCTCATTTCAACCACCCTACCCAACCCCCCACTGCCTGCCAG  
CCACCACTCCCTCCCACATTTGCAGGCGGGGGCCCTGCCCTCTCCTGCCG  
CTGGTTCCCCTACCCAGGAGGCTCTCCCATCGCTCTTTTGAGAGTCTGCC

TCCCACCTCTAACTGGGGGCTTAGTTCAAGTTGCCCCCTTACCCTAGTCC  
CAGCTGCCCCAAGAGCTTGCTGCCTCCTGTTCTTGGTGAGGGACTCCAGAG  
ACAGATGTGAGACCTCCCTGGACCCCTCCAAGGCATTCCCAGGTCACTTC  
CATGAGTAGTGAAGAACCGCCTCTGAGCAGGCTGAGCCTCCCTCAGCCTA  
TGGTGTCTCTACGTGGCTTGGCCACAGCAGGTGCTCACGCCTCCTCCTC  
AGCAGAGCCTACCATCCTCCTGCCATGCTCACCAGTCCCCATGCTGATAG  
CCATCACCAGTCCCCATGCTGATAGCCATCACCAGTCCCCATGCTGATAG  
CCACTTTCTGGATGCTCTAGGTCTGTCTGGATGACACAGTGACCACAGAG  
AAGGAGCTGGACACTGTGGAAGTGCTGAAAGCAATTCAGAAAGCCAAGGA  
GGTCAAGTCCAACTGAGCAACCCAGAGAAGAAGGTGGGTTTGTGTGGCA  
GGTGGGAGGGCAGTGGTGACAGAGCCAGCCGGGATAGGAGCCAGTTCGGGG  
GGCTTGGGCCATGGGACTGCTCAGGGCTGCCGAGTCCCAGCTGCGCCCCT  
CCCTGGCTGCATGACCTCGGGCAAGTCGCGGCCTCTCTGTTCTCTGTGGG  
GTGGGGACAGTGGTAGTTCCTGCTCTAAGGATATGATGAGACCATCTTTA  
CCACCCAGTTGGTGGGAACCGTTGCGCTCCCTCCTCACACCCCTGGCCTT  
GGGGAGCTCTGTGCTTCCTCTTCTCTCCCGGGCTGACTCAAGCACTCGTC  
CTCAGGGTGGTGAAGACTCCCGGCTCTCAGCTGCCCCCTGCATCAGACCC  
AGCAGCTCCCCTCCCACTGTGGCTCCCGCATCTGCCTCCCTGCCCCAGCC  
CATCCTCTCTAACCAAGGTAATCGTGTATGTATCTTGCTTCTAGTGGAGC  
CACACAGCCCTGCCTGGGCCCCCTGGCTGGGCTGGGGTTGGGGGAGAGGT  
GCCAGCACCTGCTTCCAACAGGGTCAGACACAGGGAGGGCAGTGCCTTCT  
GCAGGCTGGTCCTCGCGGGGGGACACATGGCAGGGGTGCCTGGCCTGATG  
CCAGCTGTTGCTTGCTTGGTGAGGACTCCCAATTGCTCTGATGCCCACAT  
CCAGCTCCTCTAGGAGACCGCAGGGTGTCTGACAGGCCCTGAGGCTGCCC  
TCTGAACAGGCTCGGGGCTGTTGGCTCATGGGACCCATTCCCTCACCGGC  
AGCACAAGCAGGTTGGCTCCTGGTTACAGGAAGCCGGGCTTGTGACTTTA



CTGTCTGGAGCCCGAATCCCTGTGCAGGGAAAAGCTTGCTTTTATCACTG  
 CCTCATCTCTGTGGGGTGACCCAGCCCCAGAACACCATGTTTGTGGGGCC  
 AAGATGGGCCATCTCTGTCCCTGTGGACCCATGGAAGACCAGGCCCATTC  
 GTCTGCCCCACTATCTTAGCGTTTTCAAAGGGCTTTCACCTCTGAACCCAG  
 GCATCCTCGGAGATGAGTGAGTGAAGCAGGTCTCATGAGCGTGTCTGCTG  
 GCCCGGCCCCCACGGAAGAGGGGAGGGTGTGCCGTCCCGAGTGGAGCCGA  
 GGCTCGGGACACGCAGGAAAGGACGCCGCCTGCCCGGGCTCCTGGAGACG  
 CAGAACTTGGTGTGAGGTCTTGGGAAAACAGTTCAACCCGATGTTTTAAG  
 AGCCAGAAAAACATTCCCACCCCTTGACCTGGTAACCCCACTGGTGGGGA  
 TTTTCTCTTAGAGGGATAAGATACCGGGAAGGGGAGGTGAAATGCTCACC  
 ACTGCCAAAACACGGGCTGCAACTGCAACATCGGAGGATGAGAGGGAGAG  
 TCGGCTGTGGTGCAGAATGCTCAGCAGCCCTCCCAGCAGGGACAGGAAGA  
 CTGGGCAGGAAGAGGGGAGAAGCATTCAAGTTAAGGCAAAGGCCCAACG  
 CAGAGCAGCACACTGAGGTCACACCTGTGAGATGTGGAAGAGAATTCCTG  
 AGCGTGGAGCGATGGGGTTAGGTGCCAGGATGATTGCCCATTTTGCTTCT  
 GTCAGACTCTTGACTAAGGATTTCTGGTTGCATTTTATTACATAAAAGCC  
 AGGGAGGTTATATCACGGTGAGAAAGCTTCCCTGACGCCGCCTCCTGTAG  
 CGCAGCCAAGCGAGCCTGTGGAGGTACCATATGACTGTAGGCCTCTGGGG  
 ACAGGGAGCTGCATCTGCTTCTCAAGGCCAGGGACACAGCCATTTCTGCC  
 AGCATCTGTTGATCAGTGAGTGAGTGAGTGGGCAGGTAGAGCAGGAGCCA  
 GTGAAGAGCAGGCCCTGGATGGGTGGGGATGCACCATGTCCCCAGGCTGC  
 AGCTGCAGGCAGCCCCCACATTGTCTGGAGAAGCCTCTGCACCAGCTCAG  
 CCCCCTCCTCACTCCCCTTGTGCCCTGGGGACACTCTGCAGAGGGGCACT  
 CTGCAGTCTGTCCCCGCCATCGCTGGACTTCTGGACATGGCCTCCAGATT  
 TGCACCTCTTAAATAAATCTGCAGTGAGTGTCTTTGTGTGCACCTCTCTT  
 TCCTTTTGGTGAGAAACAGCAAAGATCGGACCCCTAAGGACTCTCCTGAT

GTCTCCGCTCTATCCGCTGAGTGCCCTTTCTGACCACTTGTTTGTACAGG  
 CCACGGTCCAGGACGGGAGCAGATAGACTGTCCCTGTCCCTGTCCACATT  
 TCCTTGGTCCAAACAGGGCTTGTGGGAGGTAGTGGCAAAGGTGTTGGTC  
 TTTTCTCACTGATTTGGAGGCCTCCCCGTGTGTTTTTTCAGCCGCGTGT  
 TCCTGGGTCTTGCCTGGATGGACAGGGTTTTTTAGCGCGTGGGAGCAGCT  
 TTGCTGACCATGCCTGTTGCTTCCAGCCTGATTCCCGAGAAGGGAGCGTG  
 CTTGCGAAGGAACTGGCACTCGGGCCTGCCTGAAGGGGGCGCTGTCCAGA  
 CACACCCAGCCTCCCGTCGTGGCAGGCGCTGTCCGAGCCATGGATGATTG  
 TGACCAATAGGGGTGGTCGCCAGAGTTGATTGTCCAGCCAGGCCAGGGG  
 CTGAGAGGAGGCTGTGTGGAGAGGTGGTTAGGAGCCAGGGCTCGGTCAGC  
 TGAGTTCGCATGCCAGCTTCCTAGCTGTGGGACCTCAAGCAACTTGTAGC  
 CCCTCTGAAGCTGTTTTCTCAACTGTGAAGTGGACGCACCCTACTTCATT  
 GATTCTAAGAGGCACGCATTTCCACCTTGTGACTTCTCTGAAACTGAGGT  
 GCGTCTTTCAGTCAGTGGCGTCTCATAGTCGCTGTCAGCCAGCTGGTATT  
 CGAGATGGAGTCGTGGAAAACCCGTGGACACCTTCCGCTAGGACCAAGAT  
 GGCGCCACCTGCCGCATCTTAGATTTGATGAAATGTGGTAAATAACGAGA  
 GGCATGCATGAGCGAATGCTGGGGAGGCGCTTGGCACTACCCAGAGCTCC  
 ACAGAGGTGGTTCGATGAGGGCTGCCCTTTCCACATCCTTAGTAGGGGGT  
 TCAAGATGACCCAGACTGTGCCCCTGGGGAGCTTGGAGCCATGCGGGAGG  
 ATGAGCCATGTGCTGGAGGAGAACAGGGTAGGATGGTGTGGGGCTTTTGT  
 AGACTGTCTAGAGCAGAGAAGGTCTGCAGTGGAGGTGGTGTCTGAGGTGA  
 ATCTCGAAGGTGAATAGGAGTTGAACGTTAGCAGGCAGAGGGTGGATTGC  
 AGGAGAGCAGCGGCCTGGGCAGGTGCCCAGCGTGGCCCATCAGGGTGCTT  
 CATGCATGGCTGTGTGCTTGCCATCCTTCCTGCCTGCCTACCCCCTGCTG  
 CTTCGCTTCATGGGGGCGTTTGAGCTTGGGGCCACCTGCCTGCCTCGCTT  
 GTGGGCAGAGGACCCAGGCTGTGTGAGTTGTCCTGTCCCGGGGAGCAGCT



GAGCTTGTCCGGGTTCTCGACCTGTGGGGCTTCAGAGGACTTCGGGTCA  
TTTCAATGGGCTGTGGCGATGCTGGCTGTGGAGGTAGCCTAGGGCTCCTG  
TAGCCTTCAGTGAGACTGGCGGCCCGATGCCAGTGTTACCCCTGCTGGC  
GGCAGTCAGGAACATGTTACAAAGCTTTACTTCAAGTGGTCTAGAGGTG  
ATCTGAGGTGGAGTAACAGGTCCAGATAGGCTACGTTCATAAAACAGCTT  
CAGCGGGGTTTAGGAACACTGTGCATTTACGGGACGCAGTGGGTCAGAGT  
GCTGCTGTCCGTGGGAGGTGGCCCCAGGGCAGGTGAGTGGGCACGTCCTG  
TGGTAAGTGGGACTGTGGATGTGGGCTCAGGCTGGACTCAGCAGCCCTGC  
TGGATACCAAGGCCTGCAAGGGCTGGCCCCCTGGTGAATTGTCCCGTGCC  
CTGTGTATCTATGAGTCCTGCAGAGATGACAAATCAGGGGACGGGGTCAT  
GTCTAGTCACCGTCTGGGAAAATGCTCCAGGAGTGAACACATTTTCAGGCT  
CTTGATGGATGTACCTCCAAACTCTTCTCTGGATGGGTGGGCCAGCTTGC  
ATGCCTGTGCCGGCCTCTGCCCAGCGAGGTGAGGGCCAGGCCACACAGTC  
AGTCTGACTTTGGCAGAAGTTGAGAGGCAACACTTGTCTCTTGTTCAGC  
TTGCCTTTCTTTGTGTAATCTGAGAGCGAGCATTCTTTTCATGTTCTAT  
CCGCTGGCCGTTCTTCTGCGGAATGTCTGTTACGTCCTTTGCAGTCTGT  
TAATGAGGTTTCCAACCTTCCCTCATTTTTGTAATCTGTAAGAACTTTTT  
CCAGACTAGCGATATAAATCCTTGTCAAATATTGCAAACACTTTTCTCAT  
TTCATCTGGTTTTTAATCTATCCTGGTTTTTAAAAAATGTGTCTGTGGAAG  
TTAATTTTTATGTAGTCACATCTCAGTTTTTTTTCCATTGCATTTATTCT  
CAGAATGCTTCTCCCTGCCCTGAGATTAGATAAGCAGTCATTTGTTCTTT  
CTTGAGTTATTTTGAGATTTTCAGTTTTAACATTTTCTTCTATAATCCATG  
TGGCTGGGTTTTTGGGATCTGGCTAACCCCCGCCATGCCAGTAGCCTGAGG  
GGCCAGCCCCACTTGTTGAACAGCCGCTCTCCCCGCCCCACCCACCCTG  
CCTGCCTGCCACCCGCCCTGGTCTCTCCAGGAATCATGTTTCGTTTCAGGA  
GGAGGCCCTGGCCAGCAGCCTCTCGTCCACTGACAGTCTGACTCCCGAGC

83



AACTGCTGTGGGCCAGGGGTGGCCAGTCAGGTTTTTTTAAAAATCCGTT  
 CACAGAAGGCCTATAGAACTATTTCTTCCTCTAAAGAGACACAGATGAGA  
 TGGACTTTTCAATCTGTTTCCAAATTCTAATACCTAAACTCTGCTCAGCA  
 CATGTTGCCCTACACCAGGGGTGGCAAATCAAGGCCTGTGTGTGGCCCA  
 CAGCCTGGGAGCTAAGAATGACAGTTACATTCTTTTTTCTTTTTTTGAGA  
 CTGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGGCGTGTCTTGGC  
 TCACTGCAACCCCCGCTCCAGATTAATGCAATTTTCTGTCTCAGCCT  
 CAGCCTTCTGAGTAGCCCGGACCACAGGCGCACGCCACCACGCCCCAACTA  
 ATTTTTTATATTTTATAGTAGAGACAGAGATTCACCATGTGGCCTAGCTGG  
 TCTCGAACTCCTGAACTCCAGTGATCCACCAACCTCGGCTTCCTAAAGTA  
 CTGGAATTACAGGCATGAGCCACCGCGCCTGGCTAGAATAACAGTTACTT  
 TTTTTTTCTTTGAGACTGAGTCTTGCTTTGTCAACCAGGCTGGAGTGCAG  
 TGGCACGATCTCAGCTCGCTGCAACCTCCGCCTCCCGGGTTCAAGCGATT  
 CTTCTGCCTCAGCCACCCAAGGTGCCCCGCCACCACACCTGGCTAATTTTT  
 CTGTTTTTTAGTAGGGACAGGATTTCCGCATGTTGGACAGTTACATTCTTA  
 AAGGGCTGCTGAAGATCGTATGGACATGGTAGCCCATAAATCCCAAATG  
 TGTA CTCTGACCCTTTACAGAAGCTTACTAACTCCCACTCTACATGTGAG  
 GGCTGCGGTGGCCAAGAAGAGCTGGAATTTAAGTGTGAAGGTCCTAAGAC  
 CTGCCCCAGCCCACTTCCCTGCCCCGGAGGCCACCAGGGGTGACAAGTAG  
 ATTCATGCCCTGGAGTGTTCTTCTCTCCGGGGCTTATGGCAGCAACTGA  
 ATGACTTAGAAGTCCATGGGAGTGCTTTCTGTTGTGGGAACTCGTGTGGT  
 CTGGGCATAGCTGTGCCAGGCACCTATGGTCCAAGCCCCTAGAAGCATAG  
 ACTCTGACCAAATGCGGACCCAGCCTTCCAGCAGGCAGCACTGGCTCCC  
 ACCAGGGCCCTCATCCTGGGAACTGACTTGGCCATGTGGGAGGCTTGGGA  
 GACCCATGGGTTGGTTTCTCAGGGTCAGGGTGTAGCAGTGGGCTCCAGAT  
 GTGGCAGGTGGGAGGTGGGAGGGGCCCCCTCCAGCATGCCACTGACCTGG









GGAGTCCTCAGCTGAGCTGCTCACAGCTTTGAGGACCTGGGCAGTGAGGT  
CCTGAGTTGCCCTCCCTGGCCATTTGTGCTGTGTCACCACCTCCTGTGCC  
ACTTCCAGCCCCAGGTAGACCTCCCACCAACAGCCATCTCCCACCCCTCT  
CTTCCTCTCTGCCTTGAAGCATACGGATTCAATTGGTGAGCCAAGAGGGGC  
TTCCCATGTCTCCTTGTGGAAGCTGTGGGCATGTCCCTGGTATGTGCAGG  
TTGCTAGGGTGGTGGAGCTGACAGGAGGCCCCCGTCTTCAGGTTGAAAA  
CGAGGAGCTGAGGCACCTCATGTGGTCCTCGGTGGTGTCTACCAGACCC  
CAGGGCTGGAGGTGACTGCCTGCGTGCTGCTCTCCACCAAGGCTGTGTAC  
TTTGTGCTCCACGACGGCCTCCGCCGCTACTTCTCAGAGCCACTGCAGGG  
TAGGCACAGGGCCTGCTGGGGCTCAGGAGCTTGGAGTGTGTGGTTGGGGC  
AGGCCTGGGGGGTCAATTCTCTGGAGCCAGCTGTGTGGCTTCAGGCAGCAG  
TCAGCGACTTGGCTGCAGTGGGCTGAGAGTTCCTTGTCTGAGGAAGGGAG  
CTGTCATGAGGGAGGGGTCCATGGCCAGATGTGAACGCAGAATGCACTGA  
GCCAGGGCCTGGTGACTGCTTGGGAACAGCCTGTGATGAGAAGGGGTAG  
GCAGCCTTTGCCCCTGGGGCTGCACAGGAAGCCCTAGCCAGCGACCTGGT  
GACTCCCCTGAGCTGGAAGAGGCTCAGACTCCAGAGGGCATTGCCTATGG  
GGCTTTGCACGGGTGGAAGCCAGGCCAGCCAAGAGGACCTGTTCCCTGCTG  
GATGTGCTGCACACCTAGGAACCTTGTGCTTGCCTGCCACCGCCTCCCTC  
TGTCCCTTTCTCCATCACACAGATTTCTGGCATCAGAAAAACACCGACTA  
CAACAACAGCCCTTTCCACATCTCCCAGTGCTTCGTGCTAAAGCTTAGTG  
ACCTGCAGTCAGTCAATGTGGGGCTTTTCGACCAGCATTTCCGGCTGACG  
GGTGGGTGACCCTCTGTGCTTTGTCCTATTTTCGGGTGAAGGCCAGCATCA  
CCAGTGGGCTTCCACCTTCCGTACGTGGGTGGGTATCATAGACAGTTAT  
CTCTGTGCTCAAGAGCCACTTCTTACCCGGGGTGGGAGGAAGCAGCTTCA  
GGAAGTCTGAGAGAGCAGAACTCACGCTCCAGGGCTCAGAGCAGGAGGT  
AGGGTGTGCGGCAAGCGCTGGCCCGGACAGAAGCAGAGTGGGCCCTGGTC





6300124430

TGTCCCCAGTGCTGAGAGCAGAGAGAAGCTCATCTCGCTGTTGGCTCGCC  
AGTGGGAGGCCCTGTGTGGCCGTGAGCTGCCTGTCGAGCTCACCGGCTAG  
CCCAGGCCACAGCCAGCCTGTCGTGTCCAGCCTGACGCCTACTGGGGCAG  
GGCAGCAGGCTTTTGTGTTCTCTAAAAATGTTTTATCCTCCCTTTGGTAC  
CTTAATTTGACTGTCCTCGCAGAGAATGTGAACATGTGTGTGTGTTGTGT  
TAATTCTTTCTCATGTTGGGAGTGAGAATGCCGGGCCCCTCAGGGCTGTC  
GGTGTGCTGTCAGCCTCCCAACAGGTGGTACAGCCGTGCACACCAGTGTCTG  
TGTCTGCTGTTGTGGGACCGTTGTTAACACGTGACACTGTGGGTCTGACT  
TTCTCTTCTACACGTCCTTTCTGAAGTGTGAGTCCAGTCCTTTGTTGC  
TGTTGCTGTTGCTGTTGCTGTTGCTGTTGGCATCTTGCTGCTAATCCTGA  
GGCTGGTAGCAGAATGCACATTGGAAGCTCCCACCCCATATTGTTCTTCA  
AAGTGGAGGTCTCCCCTGATCCAGACAAGTGGGAGAGCCCGTGGGGGCAG  
GGGACCTGGAGCTGCCAGCACCAAGCGTGATTCTGCTGCCTGTATTCTC  
TATTCCAATAAAGCAGAGTTTGACACCGTCTGCATCTTCTAAACCAAGGG  
TCACTGGGATCGAGTCGACGGCCCTATAGTGAGTCGTATTAGAGCTCGCG  
GCCGCGAGCTCTAGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATC  
TGCAGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCTCCAC  
AGAGGTGGTTCGATGAGGGCTGCCCTTTCCACATCCTTAGTAGGGGGTTC  
AAGATGACCCAGACTGTGCCCTGGGGAGCTTGGAGCCATGCGGGAGGAT  
GAGCCATGTGCTGGAGGAGAACAGGGTAGGATGGTGTGGGGCTTTTGTAG  
ACTGTCTAGAAGCAAAGAAGGTCTGCAGTGGAGGTGGTGTCTGAGGTGAA  
TCTCGAAGGTGAATAGGAGTTGAACGTTAGCAGGCAGAGGGTGGATTGCA  
GGAGAGCAGCGGCCTGGGCAGGTGCCCAGCGTGGCCCATCAGGGTGCTTC  
ATGCATGGCTGTGTGCTTGCCATCCTTCCTGCCTGCCTACCCCTGCTGC  
TTCGCTTCATGGGGGCGTTTGAGCTTGGGCCCACCTGCCTGCCTCGCTTG  
TGGGCAGAGGACCCAAGCTGTGTGAGTTGTCCTGTCCCGGGGAGCAGCTG

9D

AACTGGTCCGGGGTCTCGAACTGTGGGGCTCAAAAGGACTCCGGGGTCAT  
TTCAGTGGGGCTGTGCCGATTCTTGGGGGCTGTTNGGAANGTAAAGGCCT  
AAAGGGGCTCCTGGTTANGGCCCTCAANTTTAANAACCTGGGGCCGGGG  
CCCGGAATTGCCCCCAANTTTGTTTCAACNCCCCTTGGCCTTNGGCNNGG  
GCAAATTTCCANGGGGAACCAATGGNNTTCCCCAAAANGGGGCCNTTT  
TAACCCNTTTCCAAANTTTGGGNCCTAAAAAAGGGTGGANTTCCTGAANG  
GG

(22) INFORMATION FOR SEQ ID NO: 22

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1070 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO: 22

VCLDDTVTTEKELDTVEVLKAIQKAKEVKSKLSNPEKKGGEDSRLSAAPCIRPSSSPPTVAPASA  
SLPQPILSNQGIMFVQEEALASSLSSTDLSLTPHQPIAQGCSDSLESIPAGQAASDDL RDVPGAV  
GGASPEHAEPVQVVP GSGQII FL PFTCIGYTATNQDFIQRLSTLIRQAIERQLPAWIEAANQRE  
EGQGEQGEEDDEEEEEEDVAENRYFEMGPPDVEEEEGGGQGEEDDEEEEDDEEAEEERLALWAL  
GADEDFLLEHIRILKVLWCFLIHVQGSIRQFAACLVLTDFGIAVFEIPHQESRGSSQHILSSLRF  
VFCFPHGDLTEFGFLMPELCLVLKVRHSENTLFIISDAANLHEFHADLRSCFAPQHMAMLCSPIL  
YGSHTSLQEFRLQLLTFYKVAGGCQERSQGCFFVYLVYSDKRMVQTAAGDYSNIEWASCTLCSA  
VRRSCCAPSEAVKSAAIPYWLLLTPQHNLVIKADFNPMNPNRGTHNCRNRNSFKLSRVPLSTVLLD  
PTRSCTQPRGAFADGHVLELLVG YRFVTAIFVLPHEKFHFLRVYNQLRASLQDLKTVVIAKTPGT  
GGSPQGSFADGQPAERRASNDQRPQEVPAEALAPAPVEVPAPAPAAASASGPAKTPAPAEASTSA  
LVPEETPVEAPAPPPAEAPAQYPSEHLIQATSEENQIPSHLPACPSLRHVASLRGSAIIELFHSS

91

[illegible]